

The opinion in support of the decision being entered today is *not* binding precedent of the Board.

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE BOARD OF PATENT APPEALS  
AND INTERFERENCES

---

*Ex parte* MARK W. DAVIS, JOHN MURPHY, and  
PAUL GRANT CARTER

---

Appeal 2007-2567  
Application 09/893,299  
Technology Center 2100

---

Decided: August 31, 2007

---

Before JOSEPH L. DIXON, HOWARD B. BLANKENSHIP, and  
ST. JOHN COURTENAY III, *Administrative Patent Judges*.

BLANKENSHIP, *Administrative Patent Judge*.

DECISION ON APPEAL

This appeal involves claims 1-12, the only claims pending in this application. We have jurisdiction under 35 U.S.C. §§ 6(b), 134(a).

## INTRODUCTION

The claims are directed to detecting duplicate documents by “triangulation”; i.e., the comparison of two documents to a reference document. (Specification ¶ 15.) Claim 1 is illustrative:

1. A method of detecting duplicates in a set of documents having associated nearest neighbor similarity scores, the method including:

for a particular document in the set of documents, selecting nearest neighbors of the particular document; and

flagging as potential duplicates the nearest neighbors of the particular document that have respective nearest neighbor similarity scores that are identical.

The Examiner relies on the following prior art references to show unpatentability:

|        |              |               |
|--------|--------------|---------------|
| Prager | 5,943,670    | Aug. 24, 1999 |
| Pugh   | 6,658,423 B1 | Dec. 2, 2003  |

The rejection as presented by the Examiner is as follows:

1. Claims 1-12 are rejected under 35 U.S.C § 103(a) as unpatentable over Prager and Pugh.

## OPINION

Prager describes prior art methods for document categorization. One common method was KNN, or K-nearest-neighbor. In KNN, the document to be categorized is treated as a query to a search engine. A hit-list of the best-matching documents to the one in question is generated. The top K documents from the list may be selected. Prager col. 1, l. 56 - col. 2, l. 41.

Pugh teaches that near-duplicate documents may be determined by a “transitive” property. That is, if document A is a near duplicate of document

B, and document B is a near-duplicate of document C, then document A is considered to be a near-duplicate of document C. Pugh notes that “near-duplicate” detection includes detection of documents that are exact duplicates. Pugh col. 7, l. 26 - col. 8, l. 27.

The Examiner finds that, based on these teachings, the artisan would have found it obvious to use the “transitive” property identified by Pugh to determine duplicate documents in the KNN method described by Prager, such that duplicates could be eliminated from categories, thus to provide more accurate search results.<sup>1</sup> The Examiner concludes that instant claim 1 is unpatentable under 35 U.S.C. § 103 for obviousness. (Answer 3-4.)

We do not find a response to the Examiner’s rejection in the briefs. Appellants’ remarks in the briefs seem based on the theory that the rejection proposes to combine one or more embodiments of the invention described by Pugh with one or more embodiments of the invention described by Prager. The rejection of claim 1 is based, instead, on showing what the artisan knew about categorizing documents and determining potential duplicates. “The use of patents as references is not limited to what the patentees describe as their own inventions or to the problems with which they are concerned. They are part of the literature of the art, relevant for all they contain.” *In re Heck*, 699 F.2d 1331, 1333, 216 USPQ 1038, 1039 (Fed. Cir. 1983) (quoting *In re Lemelson*, 397 F.2d 1006, 1009, 158 USPQ 275, 277 (CCPA 1968)).

The question of obviousness is resolved on the basis of underlying factual determinations that include (1) the scope and content of the prior art,

---

<sup>1</sup> Appellants acknowledge (Reply Br. 5) there was a well-known motivation to eliminate duplicates from a document set.

(2) any differences between the claimed subject matter and the prior art, and  
(3) the level of skill in the art. *Graham v. John Deere Co.*, 383 U.S. 1, 17-18, 148 USPQ 459, 467 (1966).

Instant claim 1 recites a first step of “for a particular document in the set of documents, selecting nearest neighbors of the particular document.”

As Prager demonstrates, and Appellants’ Specification confirms (Specification ¶ 12), nearest neighbor algorithms were in the prior art.

Claim 1 differs from the prior art in the second step, that of flagging as potential duplicates the nearest neighbors of the particular document that have respective nearest neighbor similarity scores that are identical. The claim does not specify what might be done with potential duplicates, but ends with the abstract idea of “flagging” nearest neighbors of a document.<sup>2</sup>

The details of the inventions described by Prager and Pugh that Appellants mine from the references demonstrate that the artisan was sufficiently skilled to flag potential duplicates as claimed. *See In re GPAC, Inc.*, 57 F.3d 1573, 1579, 35 USPQ2d 1116, 1121 (Fed. Cir. 1995) (USPTO board did not err in adopting the approach that the level of skill in the art was best determined by the references of record).

However, to show unpatentability of claim 1, there is no need for attempting some combination of embodiments of the inventions described

---

<sup>2</sup> We observe that *none* of claims 1 through 12 seem to define a statutory “process” under 35 U.S.C. § 101, but are drawn to the abstract idea of “flagging” documents. The claims appear to be nonstatutory for the same reasons identified in *Ex parte Bilski*, Appeal No. 2002-2257 (BPAI 2006), available at <http://www.uspto.gov/web/offices/dcom/bpai/its/fd022257.pdf> (USPTO Web site under the Board’s “Informative Opinions”). The decision in *Bilski* has been appealed to the Federal Circuit, where the case is pending review.

by the references. “An obviousness determination is not the result of a rigid formula disassociated from the consideration of the facts of a case. Indeed, the common sense of those skilled in the art demonstrates why some combinations would have been obvious where others would not.” *Leapfrog Enters., Inc. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1161, 82 USPQ2d 1687, 1690-91 (Fed. Cir. 2007) (citing *KSR Int’l Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1731, 82 USPQ2d 1385, 1389 (2007)). In view of the breadth of instant claim 1, Appellants’ contention (e.g., Appeal Br. 9) that nearest neighbour “technology” is inappropriate for handling sets of “billions of documents” is simply inapposite. The scope of Appellants’ claim 1 extends to flagging as potential duplicates exactly two “nearest neighbors.” One document plus two neighbor documents equals three documents. Three is considerably less than “billions.”<sup>3</sup> “What matters is the objective reach of the claim. If the claim extends to what is obvious, it is invalid under § 103.” *KSR*, 127 S. Ct. at 1742, 82 USPQ2d at 1397.

Instant claim 3 adds to claim 1 the recitation that the nearest neighbor similarity scores are calculated prior to duplicate detection for “a different purpose” than the duplicate detection and stored with the documents. Appellants acknowledge (Reply Br. 9) that Prager teaches using similarity scores (in the first instance) for document categorization, which is “a different purpose” than duplicate detection. Further, the ability to select the top K documents from a hit-list at least suggests that the similarity scores are “stored” with the documents. *See* Prager col. 2, ll. 17-26. Appellants’

---

<sup>3</sup> Even if Pugh’s invention were required for the combination, we observe that the reference expressly teaches that the invention may be used with different types of search engines -- i.e., not limited to handling sets of “billions” of documents. *See* Pugh col. 6, ll. 31-38.

remarks in support of patentability of claim 3 (Appeal Br. 11; Reply Br. 9-10) again ignore the rejection that has been applied in favor of the combination of Prager and Pugh proposed by Appellants. We are not persuaded of error in the Examiner's rejection.

We have considered claims 1 and 3 as representative in this appeal, based on Appellants' remarks in the Appeal Brief. *See* 37 C.F.R. § 41.37(c)(1)(vii). Being not persuaded of error in the rejection of any claim, we sustain the rejection of claims 1-12 under 35 U.S.C § 103(a) as unpatentable over Prager and Pugh.

#### CONCLUSION

In summary, the rejection of claims 1-12 under 35 U.S.C § 103(a) is affirmed.

AFFIRMED

ce

HAYNES BEFFEL & WOLFELD LLP  
P O BOX 366  
HALF MOON BAY, CA 94019